This version created 10/06/2010. Please provide comments at CyAir.net or by email to comments@cyair.net.

3. CyAir Recommendations

1. **Develop, distribute, and update guidance and best practices** that provide detailed instructions and information about how to make data systems interoperable. Laying out a clear, easy-to-follow path will increase participation and ensure consistency across the network of data systems. Initially, the guidance would provide a range of options and levels for data providers to become interoperable. After several years with the air quality data providers embracing and using the guidance, a set of best practices should be developed that codify the best methods, procedures, and instructions to make data systems interoperable.

    1.1. EPA (and other federal organizations) should fund, organize, and oversee a group/team to develop, within a focused 90-to-120-day period, current approaches and methods for networking air quality data and metadata that could be used in the near term to establish a "first phase" of an air quality cyberinfrastructure (AQ CI). This group would discuss their plans and seek open input (via a website) from a broad community for a 2-3 week period, as done by other organizations. This group would use the community's input and their experience to develop a guidance document that provides all the options, steps, instructions, and issues to become interoperable. The guidance would be vetted with the data providers and broader community (including ESIP and GEO) to ensure alignment with efforts of those groups. An example table of contents for the guidance document is provided in Appendix A.

    Time frame: Immediate
    Priority: High
    Experience: Interoperability, standards, communications

    1.2. Distribute the guidance using multiple communication channels to encourage understanding and adoption of the guidance. Some example methods of distribution include:
    - Components of the 101 course
    - Part of the ESIP meetings
    - Special section of the website devoted to guidance (see Rec. 5.1)

    Time frame: Immediate
    Priority: High
    Experience: Communications

    1.3. After 1-2 years of community implementation of the published guidance, conduct a 1-2 day workshop for data providers to gather experiences and lessons learned from groups that implemented the guidance. Use this information to update the guidance and create a suite of best practices.

Time frame: 2-3 years
Priority: Medium
Experience: Interoperability, standards, facilitation

1.4. Create a best practices reference document that addresses issues and experiences gleaned from the lessons learned workshop (Rec 1.3), incorporates elements of the initial guidance (Rec. 1.1) that have proven successful, and includes new technologies and methods. These best practices will provide an efficient and effective way for data systems to become interoperable with one another based on shared community experience and knowledge.

Time frame: 3-4 years
Priority: High
Experience: Interoperability, standards, communications

A number of important issues should be addressed in the guidance and best practices to make them useful and beneficial

- Create guidance that is practical, easy to understand, and easy to implement.
- Keep requirements in the guidance and best practices stable for several years (at least 3) so organizations have sufficient time to upgrade data systems and become interoperable, then long enough to operate and gain some of the benefits of interoperability.
- Clearly define best practices for system performance and security and authentication.
- Build upon previous and ongoing efforts within the GEO Air Quality Community of Practice, ESIP Air Quality Workgroup, and other organizations.
- Identify a 'neutral' individual or organization (one that is not affiliated with a particular system or standard) to coordinate the development of guidelines and best practices.

2. **Outreach and education.** An organized and targeted approach to outreach and education will be necessary to promote better understanding of the requirements and benefits of an operational AQ CI. This effort will include providing general and introductory information for managers and users of the AQ CI, as well as detailed instructions and examples for implementers of the AQ CI.

2.1 Create an AQ CI management briefing for senior management at agencies about the benefits/value and requirements/needs of cyberinfrastructure. This presentation should include definitions of terms, case studies showing the value of cyberinfrastructure, a roadmap to meet cyberinfrastructure goals, needs for sustaining an AQ CI, integration and relationship with GEOSS, intra- and inter-organizational challenges, etc.

a. Provide compelling examples of how cyberinfrastructure benefits air quality science and management. Solicit 3-5 air quality "problems" that require

interoperability to solve. Develop examples about how an AQ CI can help address these problems and save time, effort, and resources. Compare the 'problem-to-solution' process and expense with and without a functional AQ CI.

b. Determine where current efforts have fallen short conveying the benefits and how new examples can be effective. Review past cyberinfrastructure studies conducted by the National Research Counsel (NRC), National Science Foundation, and the Center for Technology in Government (CTG) to determine the key factors of successful programs.
c. Engage an organizational expert to determine how to demonstrate/explain the value of an AQ CI to senior management and program managers of data systems. Identify how changes in organizational structure/culture may be needed to successfully implement and sustain a cyberinfrastructure.
d. Develop guidelines for management/programs considering cyberinfrastructure projects, for example
    i. Steps to implementation
    ii. Use of guidance and examples
    iii. Community resources and collaboration
    iv. Required expertise
    v. Typical timelines and staffing requirements

Time frame: Short-term
Priority: High
Experience: Interoperability, standards, communications, organizational

2.2 Create an AQ CI 101 course to foster understanding, lower the barriers of entry/participation, and ultimately increase participation in the AQ CI. Conduct the course after the Task 1 guidance and best practices have been developed and follow up with annual upgrades and repeats of the course to build on lessons learned and community feedback. Target participants that run data systems that are critical to the air quality community and include managers and system developers.

a. Develop an outline for the AQ CI 101 course (see example in Appendix C).
b. Create an overview of the AQ CI highlighting the CyAir Vision, the overall concept and benefits of the cyberinfrastructure, and the benefits to the community as a whole, data providers, and users.
c. Assemble and organize guidance regarding standards, best practices, formats, and vocabularies, as described in Task 1 (also see Appendix A).
d. Create use cases and examples.
e. Assemble example applications and code samples into an archived library usable in the AQ CI 101 course and on the CyAir Resource Website (see Task 5).
f. Engage an organizational expert to examine the data provider and data user interviews and identify solutions to the sociological, organizational, and other barriers to sharing and exchanging data. Incorporate these insights in the course to enable attendees to meet organizational challenges.
g. Identify target audiences for the course, for example
    a. Data providers at government agencies
    b. Educational or research organizations

c. Developers of air quality software applications
d. Data users including researchers and analysts

h. Use multiple methods to publish course material, for example
    a. Webinars
    b. Online learning
    c. Podcasts
    d. Animated/voiceover presentations

a. Establish a timeline for conducting the course, e.g., initially after completion of Task 1, then repeated annually.
b. Evaluate and update the course annually to incorporate technology changes, lessons learned, additional examples, etc.

Time frame: Short-term
Priority: High
Experience: Interoperability, standards, communications, organizational, facilitation

2.3 Create and conduct AQ CI workshops to provide in-depth training and foster collaboration and mentorship among the air quality data provider community. Design the workshops for developers/implementers of the AQ CI approach, with sufficient technical details to facilitate system development. The workshops will follow the initial AQ CI 101 course and will be repeated annually. Workshop topics may include

i. Overall status of the AQ CI
ii. Progress made in selected areas, demonstrated successes
iii. Lessons learned
iv. Collaboration and sharing of code, examples, implementation, user interfaces, experiences
v. Overcoming organizational/institutional barriers
vi. AQ CI 'step-by-step'
    1. Web service implementation
    2. Exchange protocols and formats
    3. Metadata records
    4. Validation
    5. Performance
vii. Discussion of future trends and potential changes to the AQ CI

Time frame: Near-term
Priority: Medium
Experience: Interoperability, standards, software development

3. **Create a cyberinfrastructure of core air quality data systems**. Several efforts are underway to create the foundation of an AQ CI, but have not been embraced or expanded throughout EPA or the data systems within the air quality community. This recommendation focuses on a phased approach to building out the cyberinfrastructure. The phased approach would provide a solid start with a focus on making the high-value, operational data

systems interoperable (Table 1). For Phase 1, we recommend making several vital data-sharing systems interoperable to easily share data internally and externally to EPA. Then, in future phases, we recommend extending the experience, knowledge, and enthusiasm gained during Phase 1 to other data systems. In order to achieve this initial AQ CI, we recommend the following tasks

3.1. Distribute and implement the interoperability guidance developed in recommendation 1.
3.2. Establish a mandate and set aside funding for each of these data providers to participate in and develop the needed interoperability components.
3.3. Require data providers to participate in the AQ CI 101 course and contribute information, insights, and code examples for the CyAir Resource Website (Task 5).
3.4. Initiate a phased approach, first linking systems that will provide obvious high value with improved data exchange (see Table 1).
3.5. Publish progress and status updates on the CyAir Resource Website (Task 5). Allow data providers to work with the community organizer to report status to the broader communities involved in interoperability efforts.

Time frame: Immediate
Priority: High
Experience: Data providers, program managers, developers

3.6. After one year, review the status and progress of linking these data systems with the funding organizations and determine compliance with guidance and next steps. Publish status on the website.
3.7. Incorporate other operational data systems both within EPA and other agencies by inviting them to workshops and meetings, funding other data providers to participate, creating an interoperability mentor to help new agencies, etc.
3.8. Provide funding for operating these systems for several years beyond the initial development period.

Time frame: Near-term
Priority: High
Experience: Data providers, program managers, developers

System	Agency	Phase 1
AQS and AQS Data Mart	EPA	✓
AIRNow	EPA	✓
RSIG	EPA	✓
VIEWS	Colorado State University	✓
EIS & NEI	EPA	✓
DataFed	Washington University	✓

Table 1. Phased approach to create a cyberinfrastructure of core air quality data systems.

4. **Add cyberinfrastructure requirements to EPA contracts, grants, and procurements**. These requirements will help ensure consistent development and adoption of AQ CI best practices, community participation to share knowledge and leverage components, provide stability to the program, and support the long-term operation and viability of the AQ CI.

    4.1. EPA should add the following requirements to solicitations for cyberinfrastructure projects and programs
        - Develop interoperability components and systems consistent with the established guidance and best practices.
        - Create sufficient documentation to describe the data, metadata, services, etc.
        - Contribute and participate in the air quality communities by attending workshops, providing example code, and contributing to the community website.
        - Report annual status on accomplishments to the CyAir website and the broader air quality community.
        - Demonstrate a plan for long-term operation and maintenance of the system beyond the one-time development project.
        - Provide a plan for updating and improving the interoperability components in future years when a "technology refresh" is needed.
        - Establish a point of contact as a resource for work assignment managers.

    Time frame: Immediate
    Priority: High
    Experience: Program managers, contracts

    4.2. EPA should create outreach and education materials so that program managers embrace and implement these new requirements. Develop checklists and protocols for implementing these new requirements. Create outreach and communications to inform program managers about the requirements and the benefits of the requirements. Incorporate this material into the AQ CI Management Briefing (Task 2).

    Time frame: Short-term
    Priority: High
    Experience: EPA management (?)

    4.3. After several years of implementation, EPA should evaluate the compliance and effectiveness of these new requirements.

    Time frame: Near-term
    Priority: High
    Experience: Program evaluator (contractor)

5. **CyAir resource website.** A well-designed, managed, maintained, and up-to-date website with features for both publishing information and obtaining feedback will be a key resource for those developing and using the AQ CI. The website will serve as a central location for information and collaboration regarding CyAir activities and provide a resource for various communities.

   5.1 Develop and maintain a website that provides information and guidance that is practical, easy to understand, and easy to implement. Leverage and coordinate with resources on the ESIP Air Quality Workgroup and GEO Air Quality Community of Practice websites, cross-linking with these websites where appropriate. Elements of the CyAir website would include

   a. CyAir information and status
      i. The CyAir Vision
      ii. Roadmap for the future
      iii. Current status (regularly updated) including a visual diagram of the current state of the AQ CI
   b. Education and outreach materials
      i. AQ CI 101 course material and resources
      ii. AQ CI Management Briefing background material and examples
      iii. AQ CI workshop materials including use cases, code samples, and reference implementations
   c. Resources
      i. Guidance and best practice documentation
      ii. Example code
      iii. Archive of presentations, white papers, and other documents
   d. Social media tools
      i. RSS feeds
      ii. Twitter posts
   e. Directory of online resources
      i. Community websites (ESIP, GEO AQ CoP)
      ii. Air quality service catalogs, discovery tools, and documentation
      iii. Metadata record generation and validation tools
      iv. Web service validation tools
      v. Other resources useful for AQ CI
   f. A feedback/discussion mechanism to facilitate collaboration and capture user input, including
      i. General comment features for discussion of posted articles
      ii. Specific feedback implementations for gathering focused input by topic, e.g., standards and protocols, nomenclature, best practices, performance, tools and user interfaces
      iii. A gallery for AQ CI developers to post/share code samples, tools, etc.

   Time frame: Short-term
   Priority: High
   Experience: AQ CI developers, web developers

5.2 Explore options, determine location, and establish hosting/operation of the CyAir website, for example
   a. Operate on an EPA or other government agency server
   b. Operate at a third-party hosting company
   c. Use existing community web servers (e.g., ESIP, GEO AQ CoP)

Time frame: Short-term
Priority: High
Experience: Web developer

5.3 Provide long-term funding support for ongoing management, maintenance, outreach, and operation of the website, including
   a. A webmaster to perform regular updates and website maintenance
   b. Staff to address comments, recommendations, suggestions for the website
   c. Upgrades as new features are desired by the users
   d. Hosting costs
   e. Software licensing, if necessary
   f. Automated system backups

Time frame: Near-term
Priority: High
Experience: Web developer

6. **Provide cyberinfrastructure-building tools and resources for data providers and data users.** The creation of a stable and robust AQ CI needs to be fostered through tools to simplify and speed up the process for participating data providers and data users.

   6.1. EPA and other federal agencies should establish a data sharing and data access code library (see Rec 2.2 & 5.1). The openly available and community-shared library would provide a common space where AQ CI developers could share descriptions and code regarding their implementation of cyberinfrastructure components. The library should include
      - "Reference implementations" or "gold standards" that developers could follow to connect their systems to the AQ CI
      - Step-by-step descriptions coupled with code samples in multiple software languages to reduce the barriers to entry
      - Examples for retrieving and/or using data from multiple, disparate sources
      - Examples of tools, web applications, and clients that have integrated data access standards into their code
      - Case studies that provide examples of implementing cyberinfrastructure standards "end-to-end" from data sharing to data processing to data analysis to data visualization

Time frame: Immediate
Priority: High
Experience: Cyberinfrastructure developers, project managers

6.2. EPA and other federal agencies should support the development and adoption of interoperability compliance validation tools. These tools should build upon existing tools, such as Open Geospatial Consortium standards compliance testing tools and metadata validation tools, and extend these tools to accommodate the best practice conventions for the air quality community.

Time frame: Short-term
Priority: High
Experience: Cyberinfrastructure developers, project managers

6.3. EPA and other federal agencies should support the development of metadata creation tools. A core component of data sharing is common descriptions of data sets. Metadata standards exist but are not simple to implement in a consistent way across multiple organizations. Tools that assist data providers with generating standards compliant with metadata without requiring them to understand the details of the metadata standard will catapult the amount of comprehensive metadata available for an AQ CI.

Time frame: Short-term
Priority: High
Experience: Cyberinfrastructure developers, project managers

6.4. EPA and other federal agencies should support the development of data discovery and catalog tools. Current best practices in data sharing include searchable catalogs of metadata. Better tools are needed to assist data providers with making their metadata available to catalogs and data users with searching, retrieving, and using catalogs.

Time frame: Short-term
Priority: High
Experience: Cyberinfrastructure developers, project managers

6.5. EPA and other federal agencies should support the development of community feedback mechanisms and guidelines. Whether through online forums, email, surveys, or other methods, gathering feedback for users of cyberinfrastructure systems is essential for the advancement and relevance of these systems. Guidelines on how to capture feedback and reusable tools that systems can implement would improve responsiveness to users and enhance the experience of cyberinfrastructure.

Time frame: Short-term
Priority: Medium
Experience: Cyberinfrastructure developers, project managers

7. **Air quality community organizer/EPA liaison**. Create a full-time role at EPA for a data sharing and air quality leader/community organizer that understands the broad community needs as well as requirements for interoperability. This person would guide, monitor activities, motivate the community, and facilitate interactions with related communities such

as USGEO, ESIP, NASA, NOAA, and others. A similar position could benefit other agencies involved in AQ CI efforts.

A list of roles and responsibilities for this position are included in Appendix B. Some general roles and responsibility include

- Facilitate communications within EPA and other agencies and community groups
- Link groups and people that need help implementing components of the cyberinfrastructure
- Lead the effort to plan and organize conference and workshops
- Report the status of interoperability to EPA management

Several important considerations exist for this role

- Should be a long-term (more than 4-year) position since cyberinfrastructure implementation efforts typically take 2-3 years or more. In addition, the development of relationships and understanding with related communities takes time to establish.
- The person should have a solid understanding of air quality and the air quality community, as well as a working understanding of cyberinfrastructure and interoperability.

Time frame:	Immediate
Priority:	High
Experience:	Air quality, communications, facilitation

8. **<u>Develop a simple governance structure and leverage other communities</u>** to ensure that the efforts resulting from CyAir are planned, tracked, and embraced by the larger community. These elements of oversight and engagement with the broader air quality and cyberinfrastructure communities are part of ensuring the efforts from CyAir are successful. Specific recommendations include

8.1. Create a small, focused steering committee of data users, data providers, the Community Organizer, state/local representatives, and EPA (and other funding agencies) to help direct and carry out these recommendations. This steering committee would use the successful governance models developed by the VIEWS and AIRNow programs. This steering committee would work in an open environment by publishing their plans, status, actions, and outcomes on the website. Over time, when successful, this group would naturally broaden to include other data providers, stakeholders, and agencies.

Time frame:	Immediate
Priority:	High
Experience:	Community organizer, state/local representatives, EPA representatives

8.2. Work with the ESIP community to leverage the ideas, insights, and neutral forum that this Federation provides. Specifically, consider the following activities

- Review CyAir plans, actions, and status during their winter management meeting in Washington, D.C.
- Conduct technical meetings or workshops during their summer meetings to address details of developing, implementing, and maintaining an AQ CI
- Coordinate and seek guidance from the ESIP technology and standards committees
- Organize and conduct the AQ CI 101 course
- Provide a forum to discuss and quickly resolve technical issues related to implementing interoperable systems
- Identify and work with other information systems interested in becoming part of the AQ CI
- Implement AQ CI case studies that demonstrate the benefits of cyberinfrastructure
- Provide understanding of the current research and future trends that may impact the operational community
- Seek other groups, data providers, and organizations with whom to collaborate

Time frame: Near-term
Priority: High
Experience: Community organizer

8.3. Work with the GEO Air Quality Community of Practice to leverage global ideas and insights and share AQ CI experiences, lessons learned, and best practices with the global air quality community. Specifically, consider the following activities

- Convene a workshop to clearly identify and seek agreement on particular user community needs and requirements, which includes both current and anticipated future air quality needs. Ensure that a broad community is present at meetings and workshops to facilitate active discussions and a comprehensive understanding of air quality needs and requirements.
- Work through the GEOSS Architecture Implementation Pilot to illustrate case studies and success stories of AQ CI.
- Gain new insight from the experiences, best practices, and lessons learned from other AQ CI efforts.
- Report on status and progress of CyAir efforts.
- Link people involved in developing air quality interoperability components with others in the Community of Practice.

Time frame: Near-term
Priority: High
Experience: Community organizer

8.4. Work with the National Association of Clean Air Agencies (NACAA) to provide an understanding of cyberinfrastructure, its implementation status, and benefits to the state and local air quality agencies. Specifically, consider the following activities
- Work with NACAA to identity various opportunities to inform the state/local air quality agencies about efforts and benefits of the AQ CI.
- Inform both the Board of Directors and, at minimum, the following committees of the CyAir efforts: Criteria Pollutants, Emissions and Modeling, Monitoring, Public Education and Communications, and Training.

Time frame: Near-term
Priority: High
Experience: Community organizer

8.5. Work with NOAA to establish an air quality coordinator or point-of-contact within NOAA that can link NOAA's various air quality efforts. These include efforts at the Air Resources Laboratory, Earth System Research Laboratory, National Weather Service, National Environmental Satellite, Data, and Information Service (NESDIS), and Meteorological Assimilation Data Ingest System (MADIS).

Time frame: Near-term
Priority: High
Experience: Community organizer

Appendix A.

Example Table of Contents for Guidance Document on Interoperability

8. Security and authentication
    8.1. Security issues and considerations
    8.2. Methods for securing services
    8.3. Methods for authentication

9. Implementing interoperability
    9.1. Steps for creating an interoperable system
    9.2. Tips and tools to create an interoperable system

10. Considerations for implementation
    10.1. Adherence to standards
        10.1.1. What if the standard doesn't meet my needs?
        10.1.2. Anticipated changes to standards/methods
    10.2. Performance
        10.2.1. Guidelines for performance
        10.2.2. Ideas for improving performance (sub-setting)

11. Resources
    11.1. Website
    11.2. Code samples
    11.3. Other organizations

Appendix B
Roles and Responsibilities for the Community Organizer

Communicates with other organizations that maximize benefit for the air quality community.

- Seeks to (and helps other people) engage with organizations involved with interoperability systems, programs, and communities.
- Coordinates and works with the ESIP community by contributing to the air quality working group, organizing talks and participation at the summer technical and winter administrative meetings.
- Coordinates and works with USGEO and GEO Air Quality Community of Practice to set standards, communicate U.S. efforts on interoperability, and learn about other international efforts that could benefit the U.S.
- Coordinates and works with the National Association of Clean Air Agencies (NACAA) to identify state and local needs.
- Participates in setting future standards and conventions and communicates this information to the air quality community with sufficient time for implementation and adaption to these standards.

Listens to the groups/organizations running air quality data systems.

- Seeks opportunities to collaborate and utilize the capabilities of other organizations to address specific challenges affecting the air quality data provider community.
- Identifies the needs and challenges of the air quality community (both data providers and data users) and communicates this information to other communities (GEO, ESIP, and others).

Maintains and shares information about the status of AQ CI efforts.

- Maintains and updates the CyAir plan and task list.
- Maintains, updates, and recommends improvements to the website.
- Recruits volunteers to maintain and update the website.
- Tracks the progress of individual data providers toward interoperability.
- Communicates status of interoperability goals to the broader community.
- Evaluates the adoption and adherence to the suite of accepted standards, methods, and guidelines.
- Identifies the needs and challenges of the air quality community (both data providers and data users) and communicates this information to other communities (GEO, ESIP, and others).

Fosters an open, community approach to developing the AQ CI.

- Works with the board community to solve problems in an open, collaborative manner.
- Works with individuals, groups, and organizations to create a culture of understanding and inclusiveness.
- Links individuals and groups to solve problems.

- Understands and works to minimize and remove the organizational and institutional barriers that inhibit data exchange and interoperability.
- Helps EPA and other agencies align funding to support infrastructure development, promote data exchange, increase data usage by the community, and maintain the viability of the infrastructure in the long-term.
- Ensures that communications and information are easy to understand by a broad audience and that it is easy to participate in the community infrastructure.
- Takes a leadership role in the steering committee to help guide the initial implementation of the AQ CI.
- Ensures that a shared understanding exists within the community regarding the path to implement data sharing, methods to acquire and use data, and the benefits of participating in the community.
- Seeks to buffer data providers from rapid technology shifts and changes in organizational policy.

Appendix C

Example Outline for AQ CI 101 Course

1. Why an AQ CI?
    a. Examples of past and current data exchange issues/challenges
    b. Benefits of improved and broadened data exchange using the principles of AQ CI
    c. CyAir Vision
2. Overview of the AQ CI approach
3. Diagram of current and future AQ CI
4. Use cases
5. Examples of systems that are using the AQ CI approach to share data
6. Meeting organizational challenges by highlighting the long-term benefits and efficiency of the AQ CI
7. Steps to implement the AQ CI
    a. Guidance and best practices (see Appendix A)
        i. Standards, protocols, formats
        ii. Web services
        iii. Naming conventions (ontology)
        iv. Metadata standards
        v. Catalogs and discovery
        vi. Security and authentications
        vii. Examples and source code
    b. Data system requirements
    c. Validation and testing tools
    d. Metadata record generators and validators
    e. Maintenance and upgrades
    f. Network issues or requirements
8. Ongoing support
    a. Online course material
    b. Community resources and collaboration, mailing lists
    c. Follow-on training and workshops